



**Linx**

Revue des linguistes de l'université Paris X Nanterre

**68-69 | 2013**

**Corpus et apprentissage du français**

---

## **Constitution et exploitation de corpus d'interactions verbales pour le FLE : problèmes et programme**

Juliette Delahaie

---



### **Édition électronique**

URL : <http://journals.openedition.org/linx/1498>

DOI : 10.4000/linx.1498

ISSN : 2118-9692

### **Éditeur**

Presses universitaires de Paris Nanterre

### **Édition imprimée**

Date de publication : 19 novembre 2013

Pagination : 95-114

ISSN : 0246-8743

### **Référence électronique**

Juliette Delahaie, « Constitution et exploitation de corpus d'interactions verbales pour le FLE : problèmes et programme », *Linx* [En ligne], 68-69 | 2013, mis en ligne le 19 décembre 2015, consulté le 30 avril 2019. URL : <http://journals.openedition.org/linx/1498> ; DOI : 10.4000/linx.1498

---

Département de Sciences du langage, Université Paris Ouest

# Constitution et exploitation de corpus d'interactions verbales pour le FLE : problèmes et programme

*Juliette Delahaie,*

*Université Paris Ouest Nanterre La Défense, MoDyCo*

## 1. Introduction

Depuis les débuts de l'approche actionnelle et la publication du *Cadre européen commun de référence pour les langues* (CECRL), les interactions verbales sont entrées de manière officielle et explicite dans l'enseignement/apprentissage des langues, le CECRL (Conseil de l'Europe, 2001 : 18) indiquant qu'« on accorde généralement une grande importance à l'interaction dans l'usage et l'apprentissage de la langue étant donné le rôle central qu'elle joue dans la communication ».

En réalité, cette préoccupation n'est pas récente, Germain (1993 : 57) mentionne l'existence de manuels de dialogues fabriqués destinés à apprendre le français aux marchands anglais qui faisaient commerce avec la France, et ce dès la fin du XV<sup>e</sup> siècle. Plus récemment, *Un Niveau-Seuil* (Coste et al., 1976), dans le cadre de la mise en place de l'approche communicative dans l'enseignement des langues, mettait l'accent sur l'acte de langage et sur les contextes d'utilisation de la langue, et manifestait ainsi un intérêt certain pour différentes caractéristiques de l'interaction. Cependant, le CECRL inaugure une phase de réflexion explicite sur les interactions en s'appuyant sur des recherches effectuées essentiellement dans le domaine de la sociolinguistique et de la linguistique interactionniste. Ainsi, les travaux de Kerbrat-Orecchioni et Traverso sont exploités dans le cadre du premier *Référentiel pour le français niveau B2* (Beacco, 2004), Traverso ayant rédigé un chapitre des *Textes et références*. De plus, la description de la composante

pragmatique et sociolinguistique de la « compétence à communiquer langagièrement » dans le CECRL emprunte largement aux études anglo-saxonnes à propos de la notion de « *Face Threatening Act* » (Brown & Levinson, 1987) pour décrire la politesse et la coopération dans l'interaction, à propos du concept de script d'interaction (Schank & Abelson, 1977) ou des maximes conversationnelles (Grice, 1979).

Pourtant, cette intense réflexion ne s'est pas accompagnée d'un changement de taille au niveau de l'utilisation des données interactionnelles orales dans l'enseignement-apprentissage des langues. Les manuels de FLE accessibles sur le marché restent en effet très frileux quant à l'utilisation de données authentiques, même si les initiatives issues de la recherche en didactique des langues ne manquent pas. Celles-ci s'intéressent essentiellement aux modalités d'exploitation de données authentiques orales par les apprenants eux-mêmes. Ainsi, les travaux du Crapel<sup>1</sup> et les nombreuses réalisations didactiques déjà effectuées (le projet FLEURON et le manuel *Vacances cubaines* par exemple)<sup>2</sup> ont montré qu'il était possible de mettre en place une réelle démarche d'apprentissage à partir de l'observation et de l'exploitation de corpus oraux authentiques par les apprenants eux-mêmes. Les recherches sur le sujet foisonnent également dans le monde anglo-saxon, et l'on se reportera, entre autres, à Sinclair (2004), McEnery et al. (2006) ou à Chambers (2009) pour des études de cas et différentes modalités d'exploitation des corpus oraux en classe de langue.

Ces recherches sont extrêmement intéressantes et elles répondent à leur manière à la question de savoir comment utiliser des données authentiques pour la classe de langue. Nous voudrions cependant apporter une autre réponse et nous inscrire dans un champ plus restreint, en proposant une réflexion sur les modalités de constitution d'un corpus d'interactions verbales pour le FLE et sur son utilisation indirecte dans la classe de langue. On discutera d'abord de l'accessibilité et de la qualité des données d'interaction exploitables pour le FLE, pour ensuite montrer qu'il existe une voie intermédiaire entre d'un côté l'utilisation directe de documents authentiques par les apprenants de FLE, et d'un autre côté le dialogue fabriqué marqué par des « invraisemblances discursives » (Debaisieux, 2009 : 41), représentation d'un écrit oralisé qui ne renvoie ni aux usages du français parlé, ni à ceux de l'écrit. Entre ces deux extrêmes, il y a une place pour des dialogues réalistes, des constructions idéales plus vraies que nature, à condition de disposer de ressources interactionnelles adéquates et d'en faire une étude linguistique pertinente.

---

<sup>1</sup> Centre de recherches et d'applications pédagogiques en langues, désormais sous l'égide du laboratoire ATILF (UMR 7118) et reconnu comme équipe interne travaillant sur la didactique des langues et la sociolinguistique.

<sup>2</sup> Voir les *Mélanges CRAPEL*, 31 (2009).

## **2. La pertinence et l'accessibilité des données numériques pour le FLE en matière d'interactions verbales**

### **2.1. Quelles données d'interaction pour le FLE ?**

De quels genres d'interactions verbales a-t-on besoin pour l'enseignement des langues ? C'est à cette première question qu'il faudrait répondre avant de se lancer à la recherche de corpus accessibles, cependant les réponses sont plurielles et dépendent en grande partie du profil de chaque groupe d'apprenants. Les différents travaux du Conseil de l'Europe, à commencer par le CECRL (2001), permettent cependant de dresser un catalogue virtuel de genres interactionnels<sup>3</sup> dans lesquels enseignants et concepteurs de manuels pourraient puiser en fonction des contraintes spécifiques de leur situation d'enseignement.

Le CECRL (Conseil de l'Europe, 2001 : 61) fournit d'abord des informations pertinentes sur la compétence de l'apprenant en matière d'interaction orale et en fonction de son niveau :

- A1 : Peut répondre à des questions simples et en poser, réagir à des affirmations simples et en émettre dans le domaine des besoins immédiats ou sur des sujets très familiers.
- A2 : Peut communiquer dans le cadre d'une tâche simple et courante ne demandant qu'un échange d'information simple et direct sur des sujets familiers relatifs au travail et aux loisirs. Peut gérer des échanges de type social très courts mais est rarement capable de comprendre suffisamment pour alimenter volontairement la conversation.
- B1 : Peut communiquer avec une certaine aisance sur des sujets familiers habituels ou non en relation avec ses intérêts ou son domaine professionnel. Peut échanger, vérifier et confirmer des informations, faire face à des situations moins courantes et expliquer pourquoi il y a une difficulté. Peut exprimer sa pensée sur un sujet abstrait ou culturel comme un film, des livres, de la musique, etc.
- B2 : Peut utiliser la langue avec aisance, correction et efficacité dans une gamme étendue de sujets d'ordre général, éducationnel, professionnel et concernant les loisirs, en indiquant clairement les relations entre les idées.
- C1 : Peut s'exprimer avec aisance et spontanéité, presque sans effort.

---

<sup>3</sup> On parlera de « genre interactionnel » au sens de Kerbrat-Orecchioni et Traverso (2004). Le genre d'une interaction est défini essentiellement par des paramètres externes tels que les participants, le lieu de l'interaction, l'objectif, etc. Kerbrat-Orecchioni et Traverso (2004) reprennent en fait la grille SPEAKING de Hymes (1972).

- C2 : Possède une bonne maîtrise d'expressions idiomatiques et de tournures courantes, avec une conscience du sens connotatif.

Ce que l'on constate, c'est que ces descriptions de compétences générales en matière d'interaction orale ne sont pas exactement corrélées à une situation de communication particulière et qui serait propre à chaque niveau. Ainsi, l'activité « discussion informelle (entre amis) » fait l'objet d'une échelle de descripteurs à partir du niveau A2, même si ce n'est qu'au niveau B2 que l'apprenant est considéré comme capable de « suivre facilement une conversation animée entre amis » (ibid., p. 63). Ce n'est donc pas le matériau verbal qu'il faut changer en fonction du niveau de l'apprenant, il n'y a pas des genres interactionnels plus « faciles » que d'autres, cependant il est nécessaire d'adapter l'enseignement et les compétences travaillées.

Comment alors sérier les interactions susceptibles d'intéresser un apprenant ? D'abord, on peut se servir des quatre domaines d'utilisation de la langue présentés dans le CECRL (Conseil de l'Europe, 2001 : 18-19) :

- domaine public : tout ce qui relève des échanges sociaux ordinaires (relations commerciales et civiles ; services publics, activités culturelles, de loisirs dans des lieux publics, relations aux médias, etc.) ;
- domaine personnel : relations familiales, pratiques sociales individuelles ;
- domaine professionnel : tout ce qui concerne les interventions et les relations des acteurs dans l'exercice de leur activité professionnelle ;
- domaine éducationnel : l'acteur dans un contexte de formation.

Ensuite, on peut tenir compte des échelles de niveau présentées dans le CECRL pour différents types d'activités langagières, et qui donnent des pistes intéressantes pour établir une taxinomie d'interactions pour le FLE : « discussions informelles (entre amis) » (ibid., p. 63), « discussions et réunions formelles » (ibid., p. 64), « coopération à visée fonctionnelle (par exemple, réparer une voiture, discuter un document, organiser quelque chose) » (ibid., p. 65), « obtenir des biens et des services » (ibid., p. 66), « échange d'information » (ibid., p. 67) et « interviewer et être interviewé (l'entretien) » (ibid., p. 68).

À partir de tous ces éléments, on peut alors dresser une liste des genres d'interactions susceptibles d'intéresser un apprenant de FLE, associés à des activités langagières particulières<sup>4</sup> :

- conversations et échanges familiers : demander/donner des informations personnelles, savoir raconter une histoire, une anecdote ; se disputer ; discuter sur des thèmes variés (être capable d'argumenter) ;
- conversations de socialisation, de bon voisinage : se présenter, entretenir une relation de bon voisinage ;

---

<sup>4</sup> Selon Kerbrat-Orecchioni et Traverso (2004), les « activités langagières » ou discursives sont définies non pas par des paramètres externes, mais par des paramètres proprement linguistiques (particularités syntaxiques, associations régulières lexique-syntaxe, morphologie spécifique).

- conversations de service à finalité pratique : réserver un voyage, circuler dans les transports en commun, gérer une transaction dans différents types de magasins, savoir gérer des imprévus/difficultés dans ces différents types de situations (à partir du niveau B2) ;
- interactions de coopération : préparer un voyage, une soirée, une recette, une réunion de travail, etc. ;
- entretiens (médical, d'évaluation, de recrutement) ;
- échanges de nature professionnelle : séances de travail, discussions à plusieurs, négociations.

Ces différents types d'interactions sont connus des enseignants et des didacticiens, la plupart des manuels de FLE, généralistes ou non, reprennent plus ou moins cette taxinomie. Et pourtant, l'impression demeure malgré tout que peu de choses ont évolué dans les manuels et dans les classes. Les données authentiques d'interactions orales sont en effet peu exploitées, à commencer par les *Référentiels pour le français* qui se sont appuyés sur une « méthode intuitive pour la collecte des données linguistiques » (Beacco, 2004 : 11) ; de même, la plupart des manuels continue à fabriquer des interactions entières à partir de l'intuition des concepteurs. Pourquoi une telle situation perdure ? Parce que les données d'interactions orales qui sont susceptibles d'intéresser enseignants et concepteurs de manuels sont en fait difficiles d'accès ou inexistantes.

## **2.2. Des données d'interaction non accessibles ou non pertinentes pour le FLE**

La situation des corpus en France est bien connue, et Habert (2000) rappelle que les emplois majorés dans les corpus sont l'écrit et le français standard. C'est également ce qu'affirme Debaisieux (2009 : 42) : par rapport aux autres pays européens (Espagne, Portugal, Italie, Angleterre) et les Etats-Unis, « les corpus de langue parlée sont beaucoup plus restreints, ils sont souvent de taille modeste et rarement consultables par des personnes extérieures à la recherche locale ». À la suite d'Abouda et Baude (2006), on pourra dire qu'un corpus non accessible qui ne sert qu'une recherche spécifique mais dont on ne voit nulle part la trace, est un corpus « fantôme » qui n'existe pas. Le résultat pour l'enseignement du FLE est que les données actuellement accessibles en matière d'interactions verbales ne correspondent pas aux attentes d'enseignement/apprentissage exposées précédemment. Pour le montrer, on dressera une liste sommaire des corpus d'interactions verbales susceptibles d'intéresser la didactique.

Du côté de la recherche en didactique d'abord, aucun projet d'envergure n'est terminé depuis ESLO 1<sup>5</sup> (environ 4 500 000 mots) (voir Bergounioux et al., 1992), grand corpus d'entretiens sociolinguistiques réalisé à l'initiative d'enseignants britanniques de FLE dans la ville d'Orléans, et qui n'avait eu finalement que peu d'impact au niveau éditorial. À partir de ce que l'on appelle communément le corpus d'Orléans, deux

---

<sup>5</sup> Un ESLO 2 est en cours, mais qui n'a pas de projet didactique en amont, ce qui était le cas pour le premier corpus qui date des années 1960. Voir Baude et Dugua (2011).

manuels seulement ont vu le jour dans le domaine du FLE : la méthode anglaise *Les Orléanais ont la parole* (1976) et la méthode du BELC (12 cassettes, méthode non publiée). Plus généralement en FLE, les premières initiatives centrées autour de l'enseignement des compétences orales à partir de corpus se sont réalisées à l'étranger : le *français fondamental* (Gougenheim et al., 1964) a eu plus d'émules ailleurs qu'en France, et à part *De vive voix* (Mauger et al., 1975), parmi les manuels les plus récents qui se sont appuyés sur ses tables de fréquence, on citera la série *Parcours* (Arens & De Clercq, 1997) destinée à l'enseignement du FLE en Belgique néerlandophone, ainsi que le manuel de grammaire française d'Andersen & Fristrup (2005)<sup>6</sup>. De nos jours en FLE, l'intérêt pour les interactions verbales est nettement marqué dans le domaine du français sur objectifs spécifiques, puisque la collecte de données langagières adaptées à un programme précis d'enseignement fait partie intégrante de la méthodologie (voir Mangiante & Parpette, 2004). Autrement dit, les données d'interaction existent, les initiatives sont multiples, mais elles sont fragmentées et elles ne sont pas publiques.

Du côté de la linguistique, la réflexion sur les corpus de français parlé est actuellement intense, en témoigne l'initiative « corpus de référence du français » pilotée par l'*Institut de Linguistique Française*, et qui est justement la manifestation du fait qu'en France un tel corpus de référence, pour l'écrit comme pour l'oral, n'existe pas. Si le chercheur en didactique ou le concepteur de manuels cherche néanmoins des données d'interaction accessibles en format numérique, il aura l'embarras du choix pour un seul genre de l'oral, celui de l'entretien ; en effet, parmi les plus récents corpus accessibles en ligne, soit sous forme de transcription complète soit par l'intermédiaire de concordanciers, c'est ce genre qui est privilégié. On prendra pour exemple le Corpus de français parlé Parisien des années 2000 (CFPP2000 – voir Branca-Rosoff et al., 2012), le projet de Phonologie du français contemporain (PFC) ou le Corpus de référence du français parlé (CRFP) qui sont pour une bonne partie constitués d'entretiens<sup>7</sup>. En matière d'interactions verbales d'un autre type, notamment les interactions à visée transactionnelle, les données numérisées accessibles à la communauté des chercheurs sont en revanche relativement pauvres. Partant de l'inventaire des données orales réalisé par Cappeau et Seijido (qui date quand même de 2005) pour la DGLFLF, on peut répertorier les données interactionnelles suivantes, accessibles en partie en ligne à la communauté des chercheurs :

- CLAPI : une cinquantaine d'interactions dont les transcriptions sont accessibles en ligne<sup>8</sup> et qui peuvent intéresser la didactique : interactions commerciales (dans un tabac, au guichet de métro, dans un magasin de retouche), consultations (juridiques, chez un dentiste, médicales), corpus Bielefeld (série d'interactions lors d'un échange universitaire, natif/non-natif), conversations familiales ;

---

<sup>6</sup> La *Franske Grammatik* (2005) est un manuel de grammaire à destination de scolaires danois débutants en français ; il utilise les tables de fréquence du Français fondamental pour le choix des items présentés et intègre de façon très réussie nombre d'acquis de la linguistique sur le français parlé et les conversations.

<sup>7</sup> Le corpus PFC comprend également une partie « conversation libre » et un gros volet de lecture de textes ; CRFP contient une majorité de situations d'entretiens, et CFPP2000 uniquement des entretiens.

<sup>8</sup> La liste des corpus avec adresses Internet est donnée après la bibliographie.

- le corpus OTG : ensemble d'interactions à l'office de tourisme de Grenoble (26 000 mots) ;
- le site ASILA (projet de corpus en ligne, en vue notamment de la création de dialogues homme-machine) : il abrite un certain nombre de données pouvant intéresser la didactique du FLE, notamment des monologues dans un café, des interactions téléphoniques entre un client et un employé de la SNCF et d'Air France, et des interactions téléphoniques avec un agent du CIO (Centre d'information et d'orientation) ;
- le corpus ESLO 2 : ce nouveau corpus est composé d'entretiens, mais aussi d'interactions de la vie quotidienne, interactions dans les commerces (boulangerie), conversations au cours de repas, micro-trottoirs à la sortie du cinéma ;
- le corpus TCOF : téléchargeable dans son intégralité, le corpus comporte des interactions entre adultes comme des réunions et des assemblées, des conversations familiales, des interactions adulte/enfant.

En ce qui concerne les interactions verbales, la liste proposée par Cappeau et Seijido (2005) est en réalité plus longue, mais les données recensées ne sont souvent pas accessibles au public.

À la croisée des deux disciplines, on citera enfin un dernier corpus, le corpus Lancom répertorié dans la base de données Elicop, mais non mentionné par Cappeau et Seijido (2005) pour la Belgique (voir Flament-Boistrancourt, 2004). Lancom est un corpus différentiel natif/non-natif, enregistré pour moitié dans des classes d'enseignement secondaire en Belgique néerlandophone, et pour moitié en France (jeux de rôles ou interactions authentiques). Sa particularité est que, comme pour ESLO 1, il y a eu en amont une réflexion didactique : les premiers concepteurs de Lancom ont choisi des interactions importantes à maîtriser pour un apprenant de FLE, et qui seraient donc susceptibles d'intéresser les auteurs de manuels (entretien de recrutement, opinion sur un film, invitation entre amis, entretien pour un baby-sitting, jeux de rôles et interactions authentiques dans une agence de voyage, « aller chercher un ami à la gare », etc.). Cette orientation didactique a été associée dès le début à une intense recherche en linguistique centrée sur des faits de langue qui pouvaient éclairer l'enseignement, tels que l'emploi du conditionnel, ou la formulation de la question. Malheureusement, le corpus Lancom n'a pas vraiment été terminé, et on peut déplorer l'importante quantité de jeux de rôles du côté des locuteurs francophones : est-ce qu'en situation spontanée un locuteur réagirait et parlerait de la même manière ? La question reste ouverte mais elle mérite d'être soulevée.

En conclusion, il existe des données d'interactions verbales accessibles en ligne mais celles-ci ne couvrent, loin s'en faut, ni la variété des situations auxquelles est exposé l'apprenant de FLE, ni la variété des interactions que propose notre taxinomie issue des travaux du CECRL. De plus se pose un problème de taille pour l'enseignement du FLE, qui est celui de la représentativité des corpus.



### **2.3. Le problème de la représentativité du corpus**

Quelle matière linguistique choisir dans les interactions verbales pour l'enseignement/apprentissage d'une langue ? Dans quelle mesure peut-on avancer que telle ou telle manière de dire est plus fréquente qu'une autre, à partir de quelle taille d'échantillon d'une même situation aux paramètres par ailleurs variables (en fonction des locuteurs, du moment, du thème, etc.) ? Enseigner l'oral en interaction suppose en effet de pouvoir repérer les régularités (ou non) à l'œuvre dans tel type d'interaction ou d'activité discursive. Sensibiliser à la parole formulaire (le langage préfabriqué) suppose également une recherche des collocations les plus fréquentes à partir d'un corpus assez étendu et bien documenté (voir par exemple Sinclair, 1991 et plus récemment, Forsberg, 2008). En résumé, une grammaire des usages réels d'une langue s'appuie sur l'observation de situations diversifiées de communication dont les données doivent être représentatives : les structures syntaxiques et le lexique ne sont pas distribués de la même façon dans une conversation familière, un discours politique ou un écrit juridique. Pour s'en rendre compte, il faut travailler sur des échantillons et en explorer les régularités. Cela suppose un travail statistique qui ne peut pas se faire sur une ou deux interactions, mais à partir de quelle taille dispose-t-on d'un échantillon suffisant ?

La question de la « représentativité » des données est en fait cruciale en linguistique comme en didactique, elle est liée à la notion de « corpus de référence » défini par Habert (2000) comme un ensemble de « données aux conditions de production et de réception nettement définies et corrélées à leurs caractéristiques langagières. » Or pour qu'un corpus de référence mérite cette dénomination et soit « représentatif », il faut associer une taille suffisante à la diversité des usages représentés. Il y a donc aussi une exigence de panachage (des situations de communication) qui s'accompagne d'une démarche d'échantillonnage par laquelle on équilibre en taille les échantillons retenus, on essaie de ne pas « sur-représenter » des « lieux » du texte aux caractéristiques particulières. Le problème, c'est que les linguistes qui travaillent sur corpus ne sont pas d'accord sur la taille à partir de laquelle un échantillon est « représentatif » d'un certain usage de la langue (voir Biber, 1993), il n'y a pas de consensus sur la dimension de l'échantillon à partir de laquelle des régularités significatives commencent à émerger. Ainsi, à partir de combien d'interactions à l'agence de voyage – pour reprendre un exemple que nous avons traité dans notre thèse (Delahaie, 2008) – est-ce que l'on peut considérer que l'on dispose d'un échantillon « représentatif » ? On peut essayer de trouver une réponse dans les pratiques existantes. Le corpus PFC est par exemple constitué de 8 entretiens (au moins) par point d'enquête, cet ensemble d'entretiens étant alors censé être représentatif du parler de cette région. Cependant, la quantité de données considérées est fonction de l'objet de recherche, et l'on peut supposer que l'on n'aura pas besoin du même volume pour des phénomènes de phonologie ou pour des questions de régularités interactionnelles et discursives. La limite à partir de laquelle un échantillon serait suffisant n'étant cependant pas nettement déterminée, on peut aussi fixer la limite inférieure à partir de laquelle un échantillon est insuffisant pour l'analyse linguistique. Ainsi, un seul type d'interaction ne pourra pas être pertinent pour l'enseignement du FLE et la recherche de régularités grammaticales.

À défaut de taille critique, la question des métadonnées enfin est cruciale pour le FLE : pour travailler en classe, il faut pouvoir « recontextualiser » les interactions, ce que préconise la didactique des langues depuis au moins l'approche structuro-globale-audio-visuelle et l'usage de l'image pour la présentation de la situation. Cela veut dire qu'un corpus d'interactions pour le FLE doit être pourvu de métadonnées qui garantissent sa fiabilité, à l'image de ce que proposent la banque de données Valibel (voir Francard et al., 2002) et ESLO 2 (voir le site <http://eslo.huma-num.fr/>). Ces informations portent sur le contexte situationnel dans lequel se déroulent les interactions, elles comportent également des fiches d'identification relatives aux locuteurs.

Pour résumer, en l'absence de données actuellement pertinentes se pose la question de la constitution d'un corpus d'interactions pour le FLE, ce qui suppose une réflexion sur le type d'interactions à enregistrer en fonction des besoins des apprenants, la délimitation des données en extension (échantillonnage) et en intension (les métadonnées), et enfin une numérisation qui permette des recherches (collocations, arbres syntaxiques, etc.).

### **3. Des interactions verbales pour l'enseignement du FLE, pour quoi faire ?**

Que faire avec un corpus d'interactions verbales pour le FLE ? La question est d'autant plus pertinente qu'elle ne semble avoir passionné ni les linguistes, ni les chercheurs en didactique. On se souvient du projet colossal qu'a constitué le Corpus d'Orléans (1968), et qui ne s'est soldé que par deux activités éditoriales. Deux solutions au moins sont possibles. On peut d'abord opter pour la solution choisie par l'équipe du Crapel, à savoir favoriser l'exploitation de données orales (pas seulement des interactions) par l'apprenant lui-même, dans le cadre d'un apprentissage de la langue semi-autonome. On peut aussi partir de l'hypothèse qu'une analyse linguistique pourrait constituer une étape intermédiaire fructueuse pour une éventuelle pédagogisation des données d'interaction, et on donnera quelques exemples d'une telle démarche.

Au-delà de la question de la pertinence d'un certain lexique en fonction de l'interaction choisie, des données correctement échantillonnées permettent de relever les structures grammaticales les plus fréquentes et des emplois auxquels on n'aurait pas toujours pensé. Nous prendrons pour exemple des interactions dans une agence de voyage tirées du corpus Lancom. Il s'agit d'un type d'interaction finalisé à but externe (l'obtention d'un bien ou d'un service) qui est un grand classique des manuels de FLE. Sa maîtrise est recommandée dans le CECRL dès le niveau seuil B1<sup>9</sup>. Dans le cadre de notre travail de thèse (Delahaie, 2008), nous avons enregistré sept interactions dans une agence de voyage de la périphérie de Lille (environ 20 000 mots), que nous avons

---

<sup>9</sup> CECRL (Conseil de l'Europe, 2001 : 66) : « Peut se débrouiller dans la plupart des situations susceptibles de se produire en réservant un voyage auprès d'une agence [...] ».

utilisées conjointement avec les 6 interactions du même type enregistrées pour le corpus Lancom (1994, 10 000 mots). Il s'agit donc de données relativement importantes, bien documentées en ce qui concerne les métadonnées, notamment les locuteurs impliqués dans les interactions. De plus, et c'est ce qui fait la spécificité du corpus Lancom, ces interactions disposent d'un pendant « non-natif » puisqu'ont été enregistrés des jeux de rôles à l'agence de voyage joués par des élèves néerlandophones d'enseignement secondaire en Belgique flamande (Diksmuide, 2006, environ 20 000 mots). Le caractère différentiel du corpus est particulièrement intéressant pour l'enseignement des langues, il permet de repérer ce qui est saillant chez le natif par comparaison avec ce que ne fait pas le non-natif.

Ainsi, les données d'agence de voyage ont permis de mettre au jour un certain nombre de faits linguistiques intéressants pour l'enseignement de la compétence orale. On prendra ici l'exemple des marqueurs d'accord/assentiment et des marques d'introduction des thèmes dans l'interaction. Dans les deux cas, il s'agit de structures qui participent étroitement à la co-construction de l'interaction par les locuteurs en présence. En ce qui concerne d'abord les marqueurs d'accord, « exprimer son accord/désaccord » est un classique des manuels qui se réclament des approches communicative ou actionnelle. On les retrouve dans la plupart des manuels universalistes comme *Latitudes 3* (2010 : 181) : « exprimer l'accord et le désaccord », ou *Rond-Point 2* (Flumian et al., 2004 : 139) : « répondre à une question autrement que par oui ou non ». La notion d'accord est d'ailleurs mentionnée parmi les micro-fonctions à enseigner dans le cadre de la compétence pragmatique : « exprimer et découvrir des attitudes factuelles (accord/désaccord) » (Conseil de l'Europe, 2001 : 98). Cependant pour travailler cette compétence, les manuels proposent généralement des listes de marqueurs d'accord, à l'image du *Référentiel pour le français niveau B2* (Beacco, 2004). Ainsi, pour l'acte de langage « exprimer son accord » (2004, p.72 ; 74) après l'énoncé « Lyon est une grande ville », les auteurs proposent la liste suivante (non entièrement retranscrite ici) :

Effectivement/Sûrement [...] Tout à fait/Absolument. Complètement.  
Totalement/Sans réserves/...(Oui) sans aucun/le moindre doute. [...] Bien entendu.  
Tout à fait/Parfaitement/...d'accord [...].

Le *Référentiel* propose également des marqueurs spécifiques pour l'acte de langage « confirmer/démentir », avec des termes comme « oui (...). Tout à fait (...). [...] (Mais) si/ Si si /Bien sûr que si/Non. [...] ». Il est cependant difficile pour l'enseignant de savoir quel est le marqueur le plus adéquat pour tel ou tel type de situation ou d'activité langagière, il est difficile aussi de ne pas tomber dans la malédiction paradigmatique décriée depuis longtemps par Beacco (1989 : 144), c'est-à-dire la tendance à créer des listes de mots correspondant à une notion ou un acte de parole, dans lesquels enseignants et apprenants devront puiser librement. Or le choix d'une structure n'est pas à proprement parler « libre », il dépend de contraintes à la fois externes (la situation de communication) et internes (propres au fonctionnement langagier). C'est le cas pour les marqueurs d'accord. Dans les interactions à l'agence de voyage du corpus Lancom, il n'y a en fait qu'une palette assez réduite de marqueurs d'accord, et ce sont *voilà*, *d'accord* et *oui* qui sont les plus fréquents, dans des emplois où de plus ils ne sont pas toujours interchangeables. *Oui* étant le marqueur d'accord le

plus universel qui soit, ce sont *voilà* et *d'accord* qui nous intéressent. *Voilà* d'abord est la réponse privilégiée à une demande de confirmation ; acte intermédiaire entre l'assertion et la question et qui présente l'information à la fois sur le mode de l'apport et de la demande, la demande de confirmation est souvent signalée par un marqueur de recherche d'approbation comme *c'est ça* ou *hein* :

Exemple 1.

E- c'était l'hôtel (nom) hein **c'est ça hein ?**

C- **voilà** à Lesbos et il y a une semaine à Kyos

(« Lesbos », Lancom)<sup>10</sup>

En revanche, *d'accord* ne peut absolument pas fonctionner dans de tels contextes, mais il apparaît très souvent dans le troisième membre « évaluatif » de l'échange, « par lequel L1 clôt cet échange qu'il a lui-même ouvert, en signalant à L2 qu'il a bien enregistré son intervention réactive, et qu'il la juge satisfaisante » (Kerbrat-Orecchioni, 1998 : 236) :

Exemple 2.

C- là vous attendez une confirmation d'Aquatour ?

E- voilà

C- **d'accord**

(« Un père organisé », Lancom)

Le marqueur *d'accord* est donc à ranger dans la catégorie des « accusés de réception » qui valident une réponse, alors que la plupart des manuels le présentent uniquement comme la réponse privilégiée à une requête<sup>11</sup>.

L'analyse des interactions à l'agence de voyage permet donc de mettre en valeur non seulement les termes les plus fréquemment employés par les locuteurs francophones, mais aussi des valeurs d'emploi différentes de celles que l'on trouve dans les manuels. Dans le cas des marqueurs d'accord, une analyse des marqueurs dans d'autres types d'interactions du corpus Lancom (entretien d'embauche, jeu de rôles au téléphone pour recruter une baby-sitter, conversation entre amis) nous a permis, de plus, de vérifier leur extrême fréquence et leur caractère « trans-interactionnel ». Cependant ce n'est pas toujours le cas, et la généralisation d'emploi conduit parfois à faire de grosses erreurs.

Au sujet des dangers de la généralisation, on prendra pour exemple les introducteurs qui permettent de préciser le thème de la ou des propositions qui suivent ; ceux qui sont généralement étudiés par les linguistes, ce sont les « isolants » de Brunot (1922 : 30) : *en ce qui concerne*, *pour ce qui est de*, *quant à*, etc. Certes, ils ne font généralement l'objet d'aucune leçon, et pour « prendre la parole au cours d'une conversation », le *Référentiel pour le français niveau B2* (Beacco, 2004 : 115) n'y fait aucune allusion. Pourtant, ils sont très fréquents dans certains types d'interaction et fortement reliés à un script particulier. Cependant dans les interactions à l'agence de voyage, ce

<sup>10</sup> Dans ces extraits, E est l'employé et C un client.

<sup>11</sup> Voir Delahaie (2009) pour une présentation détaillée des différents emplois des marqueurs d'accord dans les interactions à l'agence de voyage.

ne sont ni les introducteurs *quant à*, ni *en ce qui concerne* qui apparaissent, mais uniquement *au niveau de*, et de manière beaucoup moins fréquente, *en termes de* :

Exemple 3.

E- alors = **au niveau formalité** vous avez c/e qu'i/l faut vous avez une carte d'identité ou un passeport?

(« Nice », Lancom)

Exemple 4.

E- euh et **en termes de tarifs** euh je pense que vous trouverez de toute façon il y a celui-là que vous aviez repéré aussi ?

(« Baléares », Lancom)

*Au niveau de* est un introducteur thématique que l'on retrouve dans un grand nombre d'interactions à script<sup>12</sup>, au cours desquelles il s'agit de dérouler certaines thématiques sans ordre particulier. Cet introducteur est ainsi très souvent employé dans les interactions dites du « baby-sitting » (prise de contact entre un parent et une baby-sitter potentielle), où il s'agit pour les deux interlocuteurs de s'informer sur les horaires, les tâches à faire, les études, etc. :

Exemple 5.

[...] mais euh: comment dire euh: le: au niveau euh repas euh: vous vous en occupez ou à ce moment là c'est moi qui: ?

(« baby-sitting », Lancom)

*Au niveau de* apparaît également de manière très fréquente dans les jeux de rôles d'entretien d'embauche :

Exemple 6.

F- [...] et: je voudrais sa savoir aussi = euh: par rapport euh: au salaire = **au niveau des horaires** = est-ce que c'est des horaires modulables est-ce que: c'est en partenariat avec un autre animateur?

(« entretien d'embauche », Lancom)

Cependant, cela ne veut pas dire que *niveau* est représentatif d'une manière courante d'introduire un thème. En effet, tout dépend du genre de l'interaction, et l'on ne trouvera pas ce type d'introducteur sur un marché pour poser une question sur le prix :

Exemple 7.

– Vous désirez, Madame ?

– Je voudrais des pommes s'il vous plaît.

– Combien ?

– Deux kilos.

– Voilà. Et avec ça ?

– Merci, c'est tout. Ça fait (coûte) combien ?/

???Niveau prix, ça donne quoi ?

– Ça fait (coûte) deux euros s'il vous plaît.

---

<sup>12</sup> Un script d'interaction est une séquence d'actions verbales et non verbales qui s'enchaînent de manière prédéterminée, dans une situation bien connue. Les conversations familières sont par exemple des interactions sans script prédéterminé.

L'introducteur thématique *au niveau de* n'est donc pas approprié à toutes les situations, et seule une analyse linguistique en fonction de différents genres d'interaction permet d'affiner sa valeur sémantique et de distinguer des emplois contextuellement pertinents.

Ce dernier exemple montre de plus que l'on ne sait pas tout et que la variation propre au français parlé est encore à explorer. Cela dit, l'intervention pédagogique n'est pas liée à l'exhaustivité scientifique et il est possible, en l'état de nos connaissances, de proposer des applications pertinentes.

#### **4. Quelle pédagogisation pour le FLE ? Des dialogues-modèles**

Bien sûr, il est toujours possible de travailler directement sur des données authentiques et d'encadrer les travaux des apprenants, mais cela ne veut pas dire qu'il faut rejeter en bloc la pratique bien établie des dialogues fabriqués. Les critiques à l'encontre de tels dialogues sont bien connues : ils ne rendent pas compte de la réalité de l'usage, ils sont fabriqués selon l'intuition des concepteurs et non à partir de recherches sur corpus, ils ne sont pas représentatifs, enfin, des différentes manières d'interagir à part sous forme de question/réponse (voir Andersen, 2006 à ce propos). Cependant, fabriquer des dialogues à partir de la recherche sur corpus permettrait de gommer un certain nombre de ces travers. Ces « dialogues-modèle » constitueraient une reconstruction idéale du dialogue authentique ; il s'agirait de simuler une réalité scientifiquement probable de la langue, à partir de laquelle il serait alors possible de fournir à l'apprenant des explications grammaticales et sociolinguistiques, un lexique – alors que beaucoup de manuels utilisent encore pour cela le Français fondamental qui date quand même des années 1950 – des structures et actes de langages adaptés.

Le dialogue-modèle a donc selon nous certains avantages. Il permet d'abord de répondre aux besoins de la communauté enseignante internationale qui ne se sent pas toujours capable d'utiliser elle-même des corpus de français parlé, car la compétence de l'enseignant ne se confond pas avec celle du linguiste. Le français courant doit donc être montré et même démystifié, de façon à aller à l'encontre d'une représentation persistante chez les apprenants, et même chez les enseignants, selon laquelle le français serait une langue plus « difficile » que l'anglais, associé qu'il est encore souvent, notamment dans les programmes scolaires à l'étranger, avec la tradition des Belles-Lettres. À travers le dialogue-modèle, il s'agit aussi de donner à l'apprenant et à l'enseignant les structures les plus fréquentes de la langue parlée dans telle ou telle situation d'interaction. À défaut de servir de « modèle », il donne au moins une idée statistiquement juste de la réalité. Enfin, ce type de dialogue peut s'avérer particulièrement pertinent au regard des préconisations qui sont faites en matière d'enseignement de la compétence pragmatique. Ainsi, si l'on en croit Kasper & Roever (2005 : 318), pour qu'un input soit acquisitionnellement pertinent pour l'apprenant, il doit être mis en valeur, en permettant par exemple d'attirer l'attention de l'apprenant sur les formes linguistiques qui indexent le contexte situationnel. Créer des dialogues-modèle, c'est justement attirer l'attention sur des formes linguistiques particulièrement pertinentes pour l'apprentissage, sur des points de langue qui aident l'apprenant, non seulement à maîtriser le lexique et les structures grammaticales propres à telle ou telle

situation, mais surtout à gérer la co-construction du discours en interaction. L'objectif du dialogue-modèle, c'est encore d'éveiller la « conscience métapragmatique » des apprenants ; selon Bardovi-Harlig (2001), cette dernière est entraînée par l'enseignant qui explique ce qui souvent reste implicite dans la langue. Or le dialogue-modèle est pour l'enseignant un matériel qui rend visible, saillant, ce fonctionnement de la langue.

On prendra pour exemple la construction d'un dialogue-modèle dans une agence de voyage. Ce que l'on veut mettre en valeur, c'est la co-construction du discours qui se fait à travers plusieurs marqueurs : les « marqueurs de discours », les introducteurs de thème tels que *au niveau de X*, et enfin les marqueurs d'accord les plus fréquents (*voilà, oui, d'accord*).

Le dialogue-modèle est construit à partir d'une analyse de fréquence des termes et structures que l'on souhaite mettre en valeur, ainsi qu'à partir de l'étude sémantique et pragmatique des emplois eux-mêmes<sup>13</sup>. De plus, l'utilisation d'un corpus différentiel a mis sur la piste des éléments saillants à mettre en valeur. On donne ci-dessous un exemple de jeu de rôles joué par des apprenants néerlandophones, tiré de Lancom :

Exemple 8.

E1- bonjour asseyez-vous

E2- bonjour

E1- qu'est-c/e que je peux faire pour vous ?

E2- euh nous voudrions euh réserver un séjour à Paris

E1- oui à Paris et ce serait pour quelle période↑

E3- le quatre cinq et six juin

E1- oui et vous êtes avec combien d/e personnes

E2- euh nous sommes avec quatre personnes deux adultes et deux adolescents un de seize ans et un de dix-huit ans

E1- hm hm

E3- on aim/erait loger dans un hôtel à quatre étoiles comm/e ici dans le dépliant avez-vous quelque chose à nous faire essayer ?

E1- euh oui le Hilton est très connu pour son service remarquable là vous avez euh deux télé un lit un grand lit confortable deux lits séparés et cetera il y a aussi un mini bar

E3- et le prix ?

E1- euh ce n'est pas très cher ça coûte cent euros par nuit donc mille deux cents euros pour trois jours

(jeux de rôles à l'agence de voyage, Lancom)

Ce qui frappe dans ce jeu de rôles, c'est la difficulté qu'ont les apprenants à co-construire le dialogue. Conformément à ce qui est enseigné en FLE, ces derniers mettent l'accent sur la transmission d'informations, c'est-à-dire sur l'emploi d'un lexique approprié à la situation ainsi que sur la construction de phrases grammaticalement correctes. L'ensemble est fort différent de ce qui se tresse entre les locuteurs natifs francophones :

---

<sup>13</sup> Voir Delahaie (2009) pour les marqueurs d'accord, Delahaie (2012) pour les marqueurs discursifs et enfin Delahaie & Flament (2013) pour les introducteurs de thème.

Exemple 9.

M- messieurs dames bonjour

E1- bonjour

E2- on recherche un séjour pour le mois d'août

M- hm hm

E2- euh les si

E1- plus les Baléares ouais

E2- ouais les Baléares l'Espagne si possible

M- oui pour combien de: de temps↓

E2- pour eu:h dix ou quinze jours

E1- entre dix et quinze jours

M- **d'accord** vous avez déjà regardé des choses repéré des choses↑ ou pas du tout↓

E1- bah on est venu eu:h i/l y a un ou deux mois et d/emi

E2- ouais

M- oui

E1- **puis après** on a laissé tomber puis on revient **voilà**

E2- **voilà**

M- **d'accord mais** vous n'avez rien vu

E2- **mais** on n'avait rien vu de spécial

E1- non non non

M- **OK** et vous êtes combien d/e personnes à partir↑

E1- deux

M- **d'accord** vous cherchez plutôt du: de l'hôtel du club ?

E1- hôtel club ouais hôtel club

M- plutôt du club hein

E2- euh tout inclus si possible

E1- ouais

M- **d'accord** et **au niveau d/e votre budget** vous vous êtes fixés une euh↑

E1- pas plus d/e mille euros pas personne pas plus de mille euros

M- **d'accord par contre** en formule all inclusive o: forcé obligatoirement ou de la pension complète ça peut aller↓

E2- non pension complète

E1- pension complète ça peut aller aussi

M- **oui ? d'accord = OK donc** vous êtes disponibles à partir de quand vous m/e dites euh ?

E2- douze août

E1- douze août

M- **OK** et retour grand maxi quand ?

E2- euh le

E1- eu:h j'ai pas d/e calendrier euh

E2- fin août

M- pardon <INT> M tend un calendrier à E1 </INT>

E1- merci = septembre bah nous on reprend le quatre septembre

M- **d'accord** = et **donc** quinze jours eu:h retour grand maxi donc euh

E1- **voilà** le deux ou **voilà** le deux c'est un samedi

M- **d'accord OK = au niveau de l'aéroport de départ** ça peut être Lille ou Bruxelles↓

E1- ouais

M- **d'accord**

(« Baléares », Lancom)



C'est bien une liste de questions que déroulent pourtant les apprenants néerlandophones, tout comme le fait cette employée de l'agence, mais il manque ce qui sert à construire l'interaction entre les locuteurs, à savoir :

- des introducteurs de thèmes : « et au niveau de votre budget vous vous êtes fixés une euh ? » vs « et le prix ? » ;
- des marqueurs de discours (MD) qui véhiculent des informations pragmatiques sur l'attitude des locuteurs, leur savoir partagé et la structure de l'échange (voir par exemple Schourup, 1999) : *par contre*, *donc*, *en fait* sont des MD qui n'apparaissent quasiment jamais chez les apprenants néerlandophones ;
- enfin, des marqueurs d'accord autres que *oui* : les apprenants néerlandophones font un usage quasi exclusif de *oui*, tandis que les francophones alternent avec *voilà*, *d'accord*, *OK* avec des valeurs d'emploi particulières.

Certes, on pourra faire plusieurs critiques à l'égard de cette manière de traiter les corpus. D'abord, il y a bien d'autres points à explorer (notamment la formulation de la question). Ensuite, comme du côté néerlandophone il s'agit de jeux de rôles joués par des apprenants de niveau intermédiaire, rien ne nous dit que ces derniers n'auraient pas fait autrement en situation spontanée et en conversation exolingue avec un locuteur natif. La comparaison entre les deux manières de dire sert moins à décrire ce que sait faire l'apprenant néerlandophone, qu'à mettre en valeur ce que fait le locuteur francophone en situation spontanée<sup>14</sup>.

Ces restrictions faites, on peut néanmoins proposer un dialogue-modèle pour l'enseignement/apprentissage, qui mette en valeur les éléments saillants récapitulés ci-dessus :

Exemple 10.

E- monsieur bonjour

C- bonjour

E- je vous écoute

C- **alors** [rattachement à la situation d'énonciation] **en fait** [complément d'information] on recherche un séjour pour le mois d'août

E- oui

C- **donc** [relance/reprise] euh pour quinze jours

E- **d'accord** [ratification], **donc** [reprise/relance] ce serait à partir de quand ?

C- à partir du quinze août

E- **OK** [ratification] et **au niveau de l'hôtel** [introduceur de thème] c'est pension complète que vous voulez, ou demi-pension ça peut aller ?

C- oui si on a de la demi-pension ça va aussi

E- **donc** [récapitulation] à partir du quinze août vous m'avez dit

C- **voilà** [confirmation]

E- **alors** [rattachement à la requête principale] j'ai l'hôtel *Armani* à vous proposer près d'Ajaccio en Corse, c'est un hôtel qui offre de belles prestations

C- oui ça pourrait aller et **donc** [relance/reprise] **au niveau des tarifs** [introduceur de thème] on arriverait à combien ?

---

<sup>14</sup> Pour d'autres travaux sur l'emploi des marqueurs de discours par des apprenants de L2, voir Diaol-Klaeger et Thörle (2013).

C- **donc** [conclusif] là on serait dans les 1000 euros par personne, pour quinze jours  
E- oui c'est pas mal en demi-pension hein c'est ça ?  
C- **voilà** [confirmation]  
E- écoutez je vais réfléchir et puis je vous donne une réponse rapidement  
C- **d'accord** [ratification] mais faites vite **quand même** [concession] parce que les hôtels sont déjà bien remplis  
E- **d'accord** [ratification] je vous remercie  
C- il n'y a pas de quoi

Il ne s'agira pas pour l'apprenant de répéter ce dialogue comme il est par exemple préconisé dans les méthodes audio-orales et encore audio-visuelles. Il doit surtout servir à l'éveil de la conscience métapragmatique par la mise en valeur d'emplois fréquents, et la possibilité par conséquent de les expliquer. Ce que l'on veut enfin souligner, c'est que ce dialogue-modèle ne doit pas imiter la réalité mais la styliser, un peu à la manière dont Diderot conçoit le vrai au théâtre :

[le vrai au théâtre, c'est] « la conformité des actions, des discours, de la figure, de la voix, du mouvement, du geste, **avec un modèle idéal imaginé par le poète, et souvent exagéré par le comédien.** [...] De là vient que le comédien dans la rue ou sur la scène sont deux personnages si différents, qu'on a peine à les reconnaître. » (*Paradoxe sur le comédien*, p. 21)

On est bien en effet dans la construction d'un modèle « idéal » à partir d'études de fréquence et d'emplois, et on est bien aussi dans l'exagération puisque sont mis en valeur certains éléments qui dans la réalité du dialogue spontané passent inaperçus, même à l'oreille du natif, même parfois aux yeux du linguiste averti.

## 5. Conclusion

Cet article a présenté une démarche de création de dialogues-modèles qui suppose une réflexion sur la constitution de corpus d'interactions pour le FLE. Construire une banque de données interactionnelles pour le FLE permettrait ainsi de contribuer à une grammaire de l'oral et des interactions, et de renouveler les contenus pédagogiques utilisés.

Il s'agirait finalement de concevoir une nouvelle linguistique appliquée ; celle-ci prendrait comme point de départ des préoccupations didactiques du point de vue du choix des interactions, de la constitution de corpus et de la sélection des faits linguistiques à étudier ; elle combinerait enfin les recherches faites dans le domaine de la linguistique interactionniste avec celles qui ont trait à la grammaire de l'oral, notamment celles qui ont été entreprises par Blanche-Benveniste et l'équipe du GARS, mais qui paradoxalement n'ont reçu que peu d'attention de la part des chercheurs en didactique<sup>15</sup>.

---

<sup>15</sup> Voir cependant les travaux de H. L. Andersen (2004) au Danemark ainsi que le livre de Weber (2013).

## Bibliographie

- ABOUDA, L. & BAUDE, O. (2006), « Constituer et exploiter un grand corpus oral : choix et enjeux théoriques. Le cas des ESLO », article en ligne, [http://icar.univ-lyon2.fr/ecole\\_thematique/idocora/documents/Abouda-Baude-ESLO.pdf](http://icar.univ-lyon2.fr/ecole_thematique/idocora/documents/Abouda-Baude-ESLO.pdf).
- ANDERSEN, H. L. (2004), « Comment utiliser les connaissances sur le français parlé dans l'enseignement du français langue étrangère », dans H. L. Andersen & Ch. Thomsen, (éds.), *Sept approches à un corpus : analyses du français parlé*, Berne, Peter Lang, p.187-214.
- ANDERSEN, H. L. & FRISTRUP, D. (2005), *Franske grammatik*, Kobenhavn, Alinea.
- ANDERSEN, H. L. (2006), « Authenticity in text book dialogue », dans H. L. Andersen & K. Lund (éds.), *Culture in Language Learning*, Aarhus, Aarhus University Press, p. 87-104.
- ARENS, C., DE CLERCQ, F. et al. (1997), *Parcours 1, Frans 1<sup>ste</sup> jaar SO*, Kapellen, Pelckmans.
- BARDOVI-HARLIG, K. (2001), « Empirical evidence of the need for instruction in pragmatics », dans K. R. Rose & G. Kasper (éds.), *Pragmatics in language teaching*, New York, Cambridge University Press, p.13-32.
- BAUDE, O. & DUGUA, C. (2011), « (Re)faire le corpus d'Orléans quarante ans après : quoi de neuf, LINGUISTE ? », *Corpus*, 10, p. 99-118.
- BEACCO, J.-C. (1989), « Un rendez-vous manqué? Théorie du discours et grammaire en didactique », *Le français dans le monde, recherches et applications*, Paris, Hachette, p.138-147.
- BEACCO, J.-C. (éd.) (2004), *Niveau B2 pour le français, un référentiel*, Paris, Didier.
- BERGOUNIOUX, G., BARADUC, J. & DUMONT, C. (1992), « L'étude socio-linguistique sur Orléans (1966-1991), 25 ans d'histoire d'un corpus », *Langue française*, 93, p.74-93.
- BIBER, D. (1993), « Representativeness in corpus design », *Literary and Linguistic Computing*, 8/4, p. 243-257.
- BIGGS, P. & DALWOOD, M. (1976), *Les Orléanais ont la parole : Teaching Guide and Tapescript*, Londres, Longman.
- BRANCA-ROSOFF, S. et al. (2012), « Discours sur la ville. Présentation du Corpus de Français Parlé Parisien des années 2000 (CFPP2000) », article en ligne, <http://cfpp2000.univ-paris3.fr/Articles.html>.
- BROWN, P. & LEVINSON, S. (1987), *Politeness: Some Universals in Language Use*, Cambridge, Cambridge University Press.
- CAPPEAU, P. & SEJIDO, M. (2005), *Inventaire des corpus oraux*, rapport en ligne, [http://www.dglflf.culture.gouv.fr/recherche/corpus\\_parole/Presentation\\_Inventaire.pdf](http://www.dglflf.culture.gouv.fr/recherche/corpus_parole/Presentation_Inventaire.pdf).
- Conseil de l'Europe (2001), *Cadre européen commun de référence pour les langues*, Strasbourg, Conseil de l'Europe.
- CHAMBERS, A. (2009), « Les corpus oraux en français langue étrangère : authenticité et pédagogie », *Mélanges CRAPEL*, 31, p. 15-33.
- COSTE, D., ROULET, E. et al. (1976), *Un niveau-seuil*, Strasbourg, Conseil de l'Europe.
- DELAHAIE, J. (2008), « Français parlés et français enseignés. Analyses linguistiques et didactiques de discours de natifs, de non-natifs et d'enseignants », thèse de doctorat, Université

- Paris Ouest Nanterre La Défense, [http://tel.archives-ouvertes.fr/docs/00/78/77/89/PDF/thA\\_se\\_J.Delahaie.pdf](http://tel.archives-ouvertes.fr/docs/00/78/77/89/PDF/thA_se_J.Delahaie.pdf).
- DELAHAIE, J. (2009), « *Oui, voilà ou d'accord ? Enseigner les marqueurs d'accord en classe de FLE* », *Revue Synergies pays scandinaves*, p.17-34.
- DELAHAIE, J. (2012), « Vers une transposition didactique de l'inventaire raisonné des marqueurs discursifs : aide méthodologique », dans D.Paillard & V.Thi Ngan (éd.), *Inventaire raisonné des marqueurs discursifs du français*, AUF/Université nationale de Hanoï, p. 205-277.
- DELAHAIE, J. & FLAMENT-BOISTRANCOURT, D. (2013) : « Questions de variations : autour de quelques introducteurs thématiques méconnus de l'oral, en termes de, au niveau de et par rapport à », dans C. Bolly & L. Degand (éds.), *Text-Structuring. Across the Line of Speech and Writing Variation (Corpora and Language in use series 2)*, Louvain-La-Neuve, Presses Universitaires de Louvain, p. 81-94.
- DIAO-KLAEGER, S. & THÖRLE, B. (2013), « Diskursmarker in L2 », dans C. Bürgel & D. Siepmann (éds.), *Sprachwissenschaft – Fremdsprachdidaktik : Neue Impulse*, Hohengehren, Schneider, p. 145-160.
- DIDEROT, D. (1773-1777, 1830), *Paradoxe sur le comédien*, Paris, A. Sautet et Cie.
- DEBASIEUX, J.-M. (2009), « Les corpus oraux : situation, exploitation linguistique, bilan et perspectives », *Mélanges Crapel*, 31, p. 36-56.
- FLAMENT-BOISTRANCOURT, D. (2004), « Quelle contribution un corpus natifs/non-natifs est-il susceptible d'apporter à la linguistique et à l'enseignement du français L2 : le cas de LANCOM », dans *Pré-textes franco-danois*, 4, Roskilde, Institut d'Etudes Culturelles et Linguistiques, p. 45-69.
- FLUMIAN, C., LABASCOULE, J. & ROYER, C. (2004), *Rond-Point 2*, Barcelone/Grenoble, Difusión/Presses Universitaires de Grenoble.
- FRANCARD, M., et al. (2002), « La banque de données Valibel : des ressources textuelles orales pour l'étude du français en Wallonie et à Bruxelles », dans C. D. Pusch & W. Raible (éds.), *Romance Corpus Linguistics: Corpora and Spoken Language*, Tübingue, Gunter Narr Verlag, p. 71-80.
- FORSBERG, F. (2008), *Le langage préfabriqué. Formes, fonctions et fréquences en français parlé L2 et L1*, Berne, Peter Lang.
- GERMAIN, C. (1993), *Evolution de l'enseignement des langues : 5000 ans d'histoire*, Paris, CLE International.
- GOUGENHEIM, G. & MICHEA, R. et al. (1964), *L'élaboration du français fondamental (1<sup>er</sup> degré) : étude sur l'établissement d'un vocabulaire et d'une grammaire de base*, Paris, Didier.
- GRICE, H.-P. (1979), « Logique et conversation », *Communications*, 30, p. 57-72.
- HABERT, B. (2000), « Des corpus représentatifs : de quoi, pour quoi, comment ? », dans M. Bilger (éd.), *Linguistique sur corpus : études et réflexions*, Perpignan, Presses Universitaires de Perpignan, p.11-58.
- KASPER, G. & ROEVER, C. (2005), « Pragmatics in second language learning », dans E. Hinkel (éd.), *Handbook of Research in SL teaching and learning*, Mahwah NJ, Erlbaum, p. 317-334.

- KERBRAT-ORECCHIONI, C., (1998/1990), *Les interactions verbales* (Tome 1), Paris, Armand Colin.
- KERBRAT-ORECCHIONI, C. & TRAVERSO, V. (2004), « Types d'interactions et genres de l'oral », *Langages*, 153, p. 41-51.
- LOISEAU, Y., COCTON, M.-N., et al. (2010), *Latitudes 3, méthode de français*, Paris, Didier.
- MAUGER, M.-T., ARGAUD, M. & NEVEU, P. (1975), *De vive voix*, Paris, Didier-Crédif.
- MANGIANTE, J.-M. & PARPETTE, C. (2004), *Le français sur objectif spécifique : de l'analyse des besoins à l'élaboration d'un cours*, Paris, Hachette FLE.
- MCENERY, T., XIAO, R. & TONO, Y. (2006), *Corpus-Based Language Studies*, Londres, Routledge.
- SCHANK, R. & ABELSON, R. (1977), *Scripts, plans, goals, and understanding: An inquiry into human knowledge structure*, Hillsdale NJ, Erlbaum.
- SCHOURUP, L. (1999), « Discourse markers », *Lingua*, 107, p. 227-265.
- SINCLAIR, J. (1991), *Corpus, Concordance, Collocation*, Oxford, Oxford University Press.
- SINCLAIR, J. (éd.) (2004), *Teaching and Learning Corpora*, Amsterdam, John Benjamins.
- WEBER, C. (2013), *Pour une didactique de l'oralité. Enseigner le français tel qu'il est parlé*, Paris, Didier.

## Corpus et bases de données

- Asila, Interaction langagière et apprentissage, base de données CNRS : <http://www.loria.fr/projets/asila/>.
- Corpus d'Orléans, dans la base de données Elicop, Université de Louvain-Leuven : <http://bach.arts.kuleuven.be/elicop/>.
- Corpus de français parlé parisien des années 2000 (CFPP2000), Université Paris 3-Sorbonne nouvelle : <http://cfpp2000.univ-paris3.fr/>.
- Corpus de langues parlées en interaction (CLAPI), Université de Lyon 2 : <http://clapi.univ-lyon2.fr/>.
- Corpus OTG (Office du tourisme de Grenoble), laboratoire Valoria : [http://www.info.univ-tours.fr/~antoine/parole\\_publicue/OTG/index.html](http://www.info.univ-tours.fr/~antoine/parole_publicue/OTG/index.html).
- Corpus de référence du français parlé, Université de Provence : <http://sites.univ-provence.fr/delic/corpus/>.
- ESLO 2, Université d'Orléans, (voir le site des ESLO : <http://eslo.huma-num.fr/>).
- Lancom, dans la base de données Elicop, Université de Louvain-Leuven : <http://bach.arts.kuleuven.be/elicop/>.
- Phonologie du français contemporain (PFC), base de données sur le français oral dans l'espace francophone : [www.projet-pfc.net](http://www.projet-pfc.net).
- Traitement des corpus oraux en français (TCOF), laboratoire ATILF : <http://www.cnrtl.fr/corpus/tcof/>.